TESTABLE *and* REUSABLE DATA PROCESSING

FLORIAN ECKERSTORFER
https://florian.ec

I DEVELOPED *a* LIBRARY *to* PROCESS DATA

https://florian.ec

PLUM³

A data processing pipeline for PHP.

https://florian.ec

"Data processing is the collection and manipulation of items of data to produce meaningful information."

–CARL FRENCH[1]

INFORMATION *is* DATA *with* MEANING

https://florian.ec

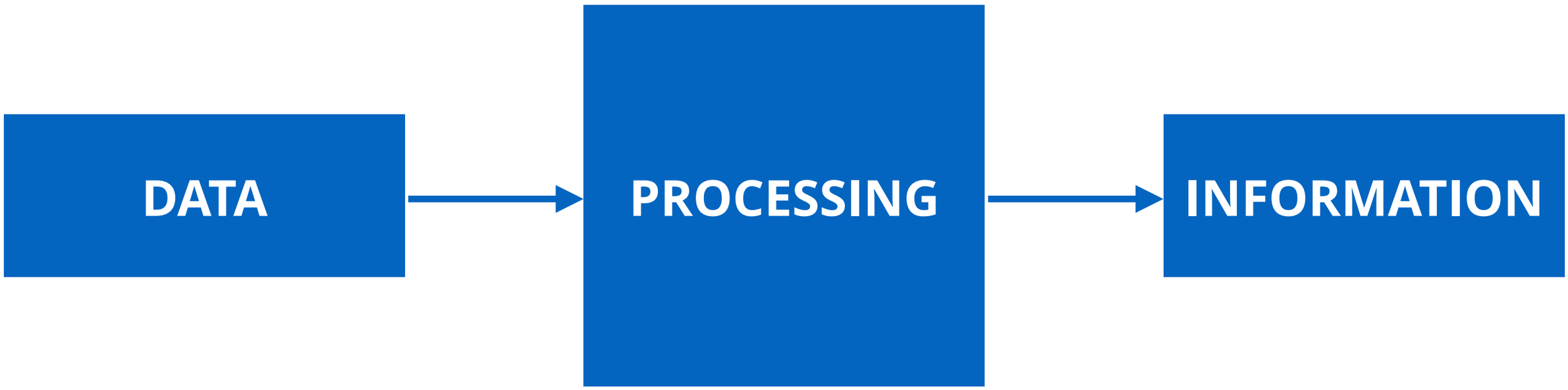
MEANING DEPENDS *on* CONTEXT

https://florian.ec

CURRENT DATE *and* TIME

CONTEXT	DATE
Newspaper	Tuesday 10 February 2015 21.42 GMT
MySQL	2015-02-10 21:42:00
PHP `date()` function	1423600920
ISO 8601	2015-02-10T21:42:00Z

https://florian.ec



https://florian.ec

PROCESSING DATA

FILTER

CONVERSION

NORMALIZATION

MAPPING

GROUPING

SORTING

https://florian.ec

The whole point of this:

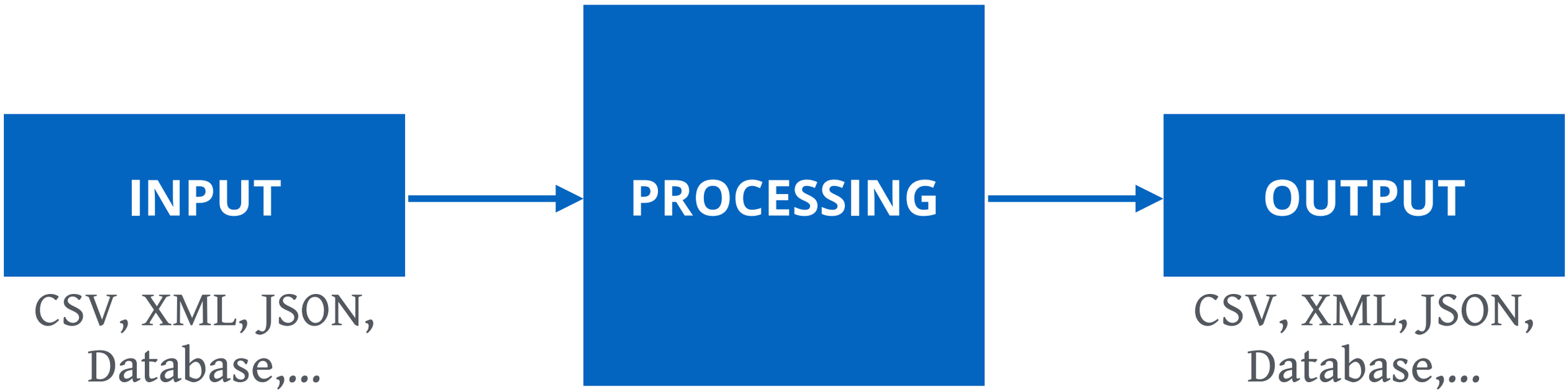
https://florian.ec

Data processing is basically all a programmer every does.

https://florian.ec

IMPORTING *and* EXPORTING DATA

https://florian.ec


INPUT
PROCESSING
OUTPUT
CSV, XML, JSON, Database,...
CSV, XML, JSON, Database,...

GitHub, Inc.

This repository Search

Explore Gist Blog Help

florianeckerstorfer

ddeboer / data-import

Watch 19 Unstar 245

Import data from and export data to a range of different file formats and media

290 commits 3 branches 21 releases 30 contributors

branch: master data-import / +

Merge pull request #157 from sagikazarmark/feature/remove_count ...

ddeboer authored 7 days ago latest commit 8f7f22561f

src/Ddeboer/DataImport	Adds CountableReaderInterface	12 days ago
tests	Add specific setSQLParameters method	12 days ago
.gitignore	Added .gitignore for lock and vendors.	a year ago
.travis.yml	Configure Travis for external code coverage	a year ago
LICENSE	Update LICENSE	a year ago
README.md	Add ExcelReader explanation (fix #147)	19 days ago
composer.json	Updated psr/log requirement to be ~1.0	a month ago
phpunit.xml.dist	Correctly ignore interfaces	3 years ago

Code

Issues

Pull Reque

Wiki

Pulse

Graphs

SSH clone URL

git@github.com

You can clone with or Subversion.

Clone in

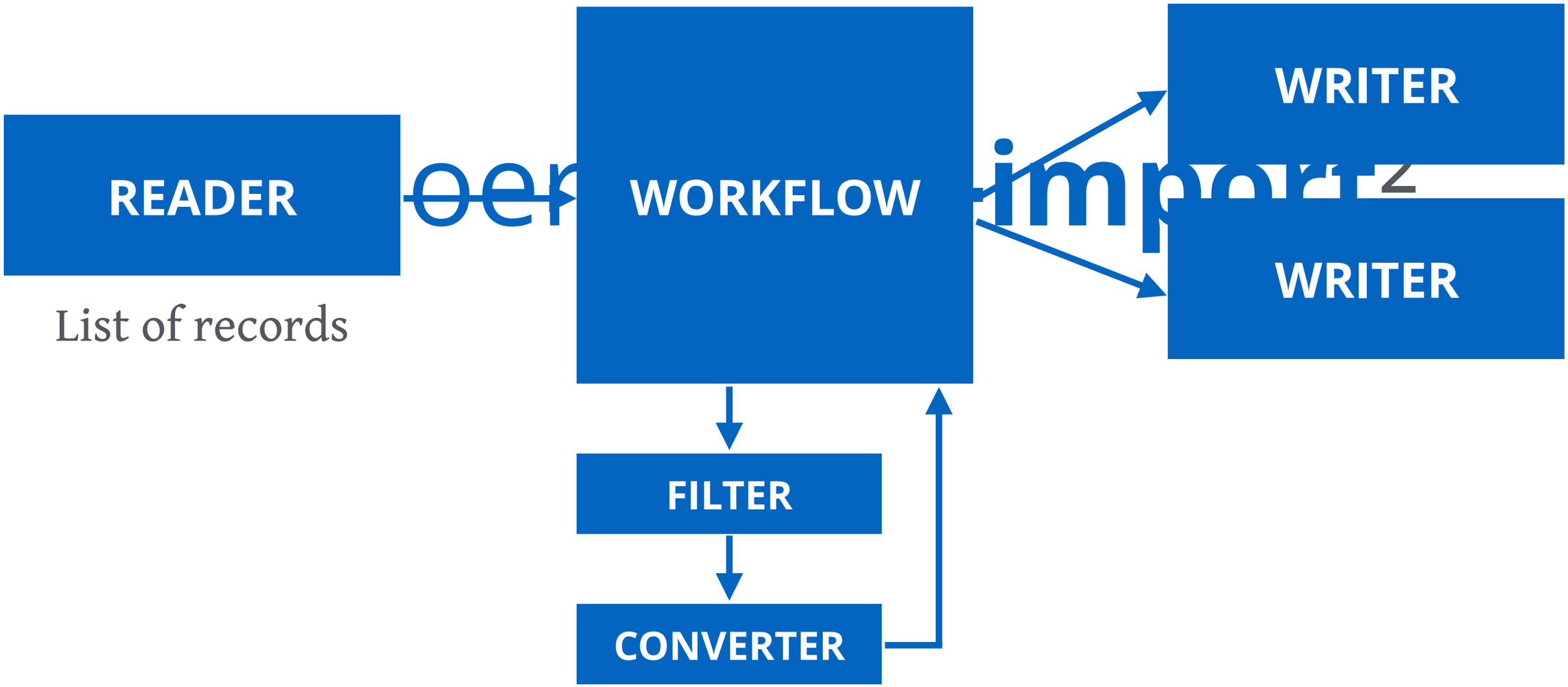
Downl

README.md

Ddeboer Data Import library

build passing quality 9.29 coverage 92 % stable 0.17.0

https://florian.ec



List of records

https://florian.ec

```php
use Ddeboer\DataImport\Workflow;
use Ddeboer\DataImport\Reader;
use Ddeboer\DataImport\Writer;
use Ddeboer\DataImport\Filter\CallbackFilter;
use Ddeboer\DataImport\ItemConverter\CallbackConverter;

$reader = new Reader\...;
$workflow = new Workflow($reader, $logger);
$result = $workflow
    ->addWriter(new Writer\...())
    ->addFilter(new CallbackFilter(...))
    ->addItemConverter(new CallbackConverter(...))
    ->setSkipItemOnFailure(true)
    ->process()
;
```

READER

```
class ArrayReader extends \ArrayIterator implements
CountableReaderInterface {
}
```

https://florian.ec

RECORD

```
[
  'name'=>'Florian',
  'age'=>28,
  'birthday'=>DateTime object(…)
]
```

FILTER

```php
use Ddeboer\DataImport\Filter\ValidatorFilter;

$filter = new ValidatorFilter($validator);
$filter->add('email', new Assert\Email());
$filter->add('sku', new Assert\NotBlank());
```

CUSTOM FILTER

```php
interface FilterInterface
{
    public function filter(array $item);
    public function getPriority();
}
```

https://florian.ec

CONVERT

```php
use Ddeboer\DataImport\ValueConverter\DateTimeValueConverter;

$converter = new DateTimeValueConverter('d/m/Y H:i:s', 'd-M-Y');
$workflow->addValueConverter('birthday', $converter);
```

CUSTOM CONVERTER

```php
interface ItemConverterInterface
{
    public function convert($input);
}
```

```php
interface ValueConverterInterface
{
    public function convert($input);
}
```

https://florian.ec

WRITER

```php
interface WriterInterface
{
    public function prepare();
    public function writeItem(array $item);
    public function finish();
}
```

SMALL COMPONENTS

Reader

Writer

Filter

Item Converter

Value Converter

https://florian.ec

EACH COMPONENT

Easily testable

Reusable with other workflows

Reusable without workflow

https://florian.ec

So what's the problem? Why did you create your own data processing library?

https://florian.ec

DATA TYPE *of* RECORD

data-import

Array

Plum

Array

String

Object

Integer

https://florian.ec

```php
$files = [
    './file1.md',
    './file2.md',
    './file3.md'
];
```

```php
class ReadFileConverter implements ConverterInterface {
    public function convert($item) {
        return ['file'=>$item, 'content'=>file_get_contents($item)];
    }
}
```

```php
$files = [
    ['file'=>'./file1.md','content'=>'Hello *World*!'],
    ['file'=>'./file2.md','content'=>'**Foobar**']
    ['file'=>'./file3.md','content'=>'![](img1.png)']
];
```

```php
class FinderReader implements ReaderInterface
{

    private $finder;

    public function __construct(Symfony\Component\Finder\Finder $finder)
    {
        $this->finder = $finder;

    }

    public function getIterator()
    {
        return $this->finder->getIterator();
    }

    public function count()
    {
        return $this->finder->count();
    }
}
```

ORDER *of* PROCESSING

data-import

Plum

1. Filter

2. Item Converter

3. Value Converter

4. Filter

5. Writer

https://florian.ec

```php
interface FilterInterface
{
    public function filter($item);
}

interface WriterInterface
{
    public function writeItem($item);
}

interface ConverterInterface
{
    public function convert($item);
}
```

```php
interface FilterInterface extends PipelineInterface
{
    public function filter($item);
}

interface WriterInterface extends PipelineInterface
{
    public function writeItem($item);
}

interface ConverterInterface extends PipelineInterface
{
    public function convert($item);
}
```

```
$workflow = new Workflow();
$workflow->addFilter($filter)
         ->addConverter($dateConverter)
         ->addWriter($csvWriter)
         ->addConverter($textEncodingConverter)
         ->addWriter($xmlWriter);
$workflow->process($reader);
```

ORDER *of* PROCESSING

data-import

1. Filter
2. Item Converter
3. Value Converter
4. Filter
5. Writer

Plum

Filter

Converter

Writer

https://florian.ec

ORDER *of* PROCESSING

data-import

1. Filter
2. Item Converter
3. Value Converter
4. Filter
5. Writer

Plum

Filter

Converter

Filter

Writer

ORDER *of* PROCESSING

data-import	Plum
1. Filter	Filter
2. Item Converter	Writer
3. Value Converter	Filter
4. Filter	Writer
5. Writer	

ORDER *of* PROCESSING

data-import	Plum
1. Filter	Converter
2. Item Converter	Writer
3. Value Converter	Filter
4. Filter	Converter
5. Writer	Writer

PLUM

A data processing pipeline for PHP.

https://florian.ec

Arbitrary data type of record

Arbitrary order of processing

Works with Dependency Injection

No value converters

https://florian.ec

WORKFLOW CONCATENATION

Reuse workflows

Useful with Dependency Injection

```
$concatenator = new WorkflowConcatenator();

$workflow1->addWriter($concatenator);
$workflow1->process($reader);

$workflow2->process($concatenator):
```

https://florian.ec

CURRENT STATUS

Workflow + Interfaces

= Design pattern implemented

https://florian.ec

IN PROGRESS

Modular system

plumphp / **plum**

plumphp / **plum-json**

plumphp / **plum-csv**

plumphp / **plum-date**

https://florian.ec

ROADMAP

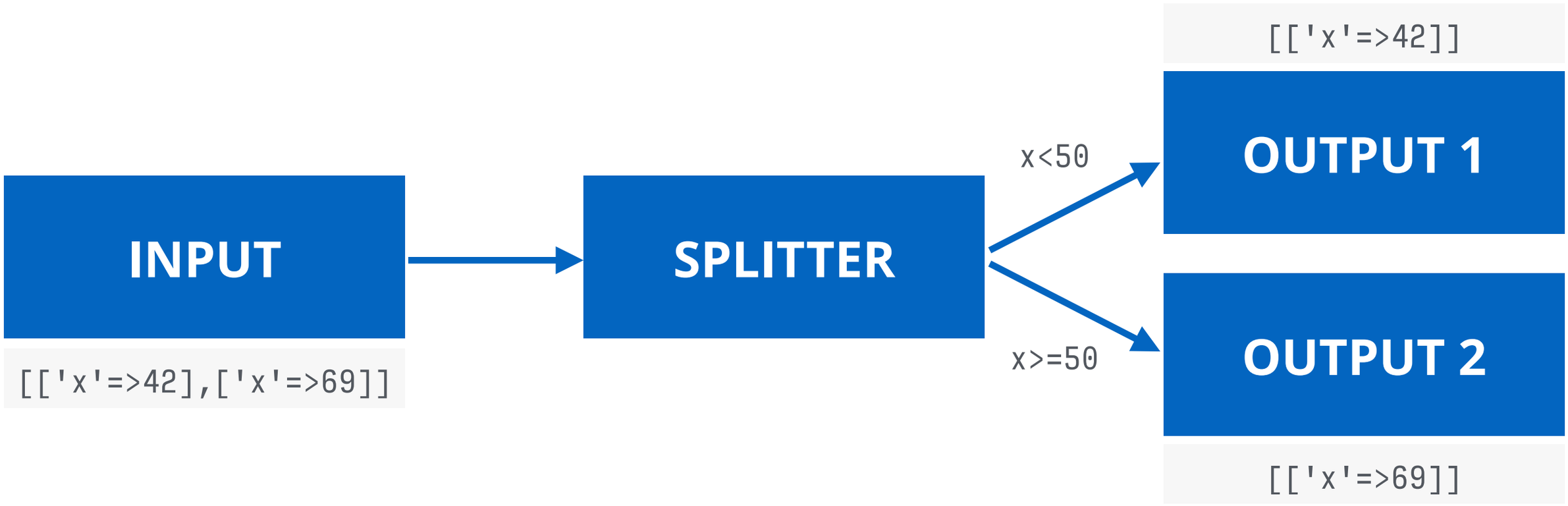
Workflow Splitter

Value Converters

Logging

https://florian.ec

WORKFLOW SPLITTER



https://florian.ec

VALUE CONVERTER

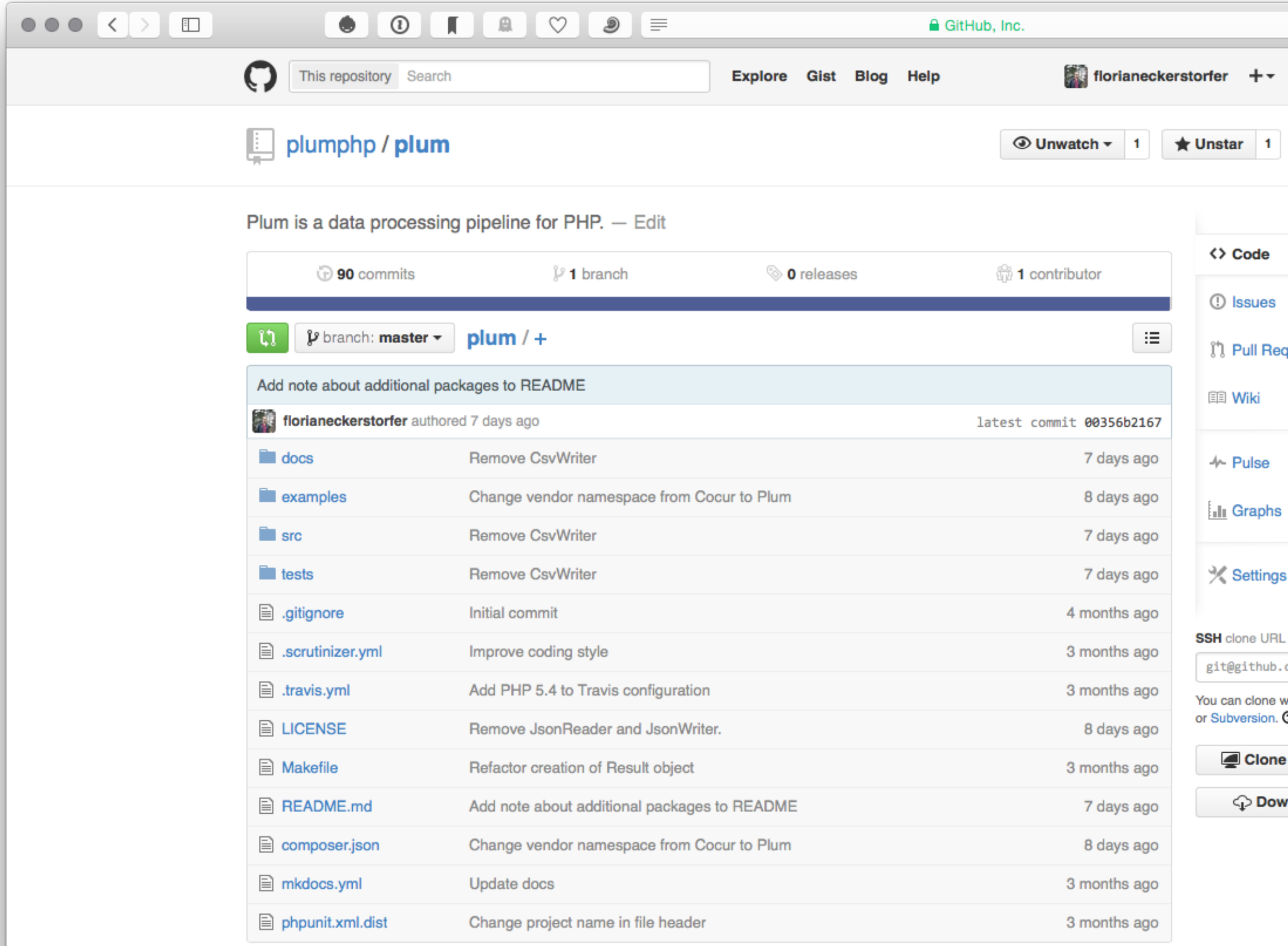
Symfony Property Access Component

```php
$converter = new UppercaseConverter('[name]');
$converter->convert(['name' => 'florian']);
```

```php
$converter = new UppercaseConverter('name');
```

```php
$converter = new UppercaseConverter('children[0].name');
```

https://florian.ec


GitHub, Inc.
This repository
Search
Explore
Gist
Blog
Help
florianeckerstorfer
plumphp / plum
Unwatch
1
Unstar
1
Plum is a data processing pipeline for PHP. — Edit
90 commits
1 branch
0 releases
1 contributor
branch: master
plum / +
Add note about additional packages to README
florianeckerstorfer authored 7 days ago
latest commit 00356b2167
docs Remove CsvWriter 7 days ago
examples Change vendor namespace from Cocur to Plum 8 days ago
src Remove CsvWriter 7 days ago
tests Remove CsvWriter 7 days ago
.gitignore Initial commit 4 months ago
.scrutinizer.yml Improve coding style 3 months ago
.travis.yml Add PHP 5.4 to Travis configuration 3 months ago
LICENSE Remove JsonReader and JsonWriter. 8 days ago
Makefile Refactor creation of Result object 3 months ago
README.md Add note about additional packages to README 7 days ago
composer.json Change vendor namespace from Cocur to Plum 8 days ago
mkdocs.yml Update docs 3 months ago
phpunit.xml.dist Change project name in file header 3 months ago
Code
Issues
Pull Req
Wiki
Pulse
Graphs
Settings
SSH clone URL
git@github.
You can clone w
or Subversion.
Clone
Dow

You can contribute on Github

Feedback

Bugs

Pull Requests

https://florian.ec

DO YOU HAVE ANY QUESTIONS?

or FEEDBACK?

https://florian.ec

[1] French, Carl (1996). Data Processing and Information Technology (10th ed.).

[2] David de Boer. data-import. https://github.com/ddeboer/data-import

[3] Florian Eckerstorfer. plum. https://github.com/plumphp/plum